DOCUMENT RESUME

ED 064 395 TM 001 643

AUTHOR	Byars, Jackson A.; Roscoe, John T.
TITLE	Rational Approximations of the Inverse Gaussian Function.
PUB DATE	Apr 72
NOTE	7p.; Paper presented at the annual convention of the AERA (Chicago, Ill., April 1972)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Behavioral Objectives; *Comparative Analysis; *Mathematical Models; *Measurement Instruments; Numbers; Standards; *Statistical Analysis; Symbols (Mathematics)

ABSTRACT

There are at least two situations in which the behavioral scientist wishes to transform uniformly distributed data into normally distributed data: (1) In studies of sampling distributions where uniformly distributed pseudo-random numbers are generated by a computer but normally distributed numbers are desired; and (2) In measurement applications where standardization of an instrument requires that percentile ranks be transformed into normally distributed standard scores. The problem investigaged in this study is find z when given P(z). The difficulty is that expressions which approximate the integral from minus infinity to z are not readily solvable for z. A number of investigators have derived algebraic approximations to the inverse Gaussian. The most widely used algebraic approximations of the inverse Gaussian function are those derived by Hastings. The Hastings approximations are valid only for values of P(z) greater than 0.50, and a computer program must make logical provisions for the situation where P(z)0.50. Burr approached the problem through the use of a cumulative moment theory and also derived two approximations. Burr's approximations have the advantage that they are valid for all values of P0. They are also conveniently expressed in one FORTRAN statement. It was the objective of Byars and Roscoe to develop an approximation of the inverse Gaussian which was both more accurate and more efficient than previous transformations. A final expression was obtained from the solution of approximately 4300 equations in six unknowns. The three sets of approximations were compared on accuracy and computational efficiency, and the Byars-Roscoe approximation was found to be superior to the others. (CK)

3.10

ED 064395

RATIONAL APPROXIMATIONS OF THE INVERSE GAUSSIAN FUNCTION[1]

Jackson A. Byars and John T. Roscoe, Kansas State University

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

BACKGROUND

There are at least two situations in which the behavioral scientist wishes to transform uniformly distributed data into normally distributed data:

(1) In studies of sampling distributions where uniformally distributed pseudo-random numbers are generated by a computer but normally distributed numbers are desired. Such applications frequently involve Monte-Carlo studies in which very large samples of data are drawn. In such cases computational efficiency becomes a prime criterion for normalization functions.

(2) In measurement applications where standardization of an instrument requires that percentile ranks be transformed into normally distributed standard scores. In such applications accuracy as well as computational efficiency is of importance.

In either situation it would be desirable to have an efficient, accurate procedure for use in computer programs.

The standard normal cumulative distribution function is given by:

$$P(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$$

The problem investigated in this study is to find z given P(z). The difficulty is that expressions which approximate the integral from minus infinity to z are not readily solvable for z. It is possible to obtain a value for z which is as accurate as may be desired by means of making successive approximations of the value of the integral for various values of z. Such procedures are computationally inefficient and are not considered in this paper.

TM 001 543

---

[1]Paper presented at the annual convention of the American Educational Research Association, Chicago, April. 1972.

ERIC

A number of investigators have derived algebraic approximations to the inverse Gaussian. These approximations are of varying degrees of accuracy and computational efficiency. In this paper an algebraic expression is presented which is more accurate than previous expressions over the range $0.01 \leq P(z) \leq 0.99$ and is more computationally efficient.

HASTINGS' APPROXIMATIONS

The most widely used algebraic approximations of the inverse Gaussian function are those derived by Hastings (1). He used Chebyshev polynomials to derive two approximations, the first simpler than the second but yielding a less accurate approximation. The Hastings approximations are valid only for values of P(z) greater than 0.50, and a computer program must make logical provisions for the situation where $P(z) \geq 0.50$. Since the approximations are found on sheets 67 and 68 respectively of Hastings' book, they are referred to in this paper as Hastings(67) and Hastings(68) respectively.

Hastings(67)

for $P(z) < 0.50$

$$n = \sqrt{\ln \frac{1}{(1 - P(z))^2}}$$

$$z = n - \left\{ \frac{a_0 + a_1 n}{1 + b_1 n + b_2 n^2} \right\}$$

$a_o = 2.30753$ $\qquad b_1 = 0.99229$

$a_1 = 0.27061$ $\qquad b_2 = 0.04481$

Hastings(68) with P(z) and n as above

$$z = n - \left\{ \frac{a_0 + a_1 n + a_2 n^2}{1 + b_1 n + b_2 n^2 + b_3 n^3} \right\}$$

$a_0 = 2.515517$ $\qquad b_1 = 1.432788$

$a_1 = 0.802853$ $\qquad b_2 = 0.189269$

$a_2 = 0.010328$ $\qquad b_3 = 0.001308$

For purposes of this investigation, the above formalae were written in FORTRAN IV as follows:

```
IF(P.EQ.0.500) P = 0,50000001
C = ABS(P-.5)
T = SQRT (ALOG(1./(.5-C)**2))
D = (P-.5)/C
H67 = D* (T-((2.30753+0.27061*T)/(1.+(0.99229+0.04481*T)*T)))

H68 = D*(T-((2.515517+(.802853+.010328*T)*T)/(1.+(1.432788+
&(.189269+.001308*T)*T)*T)))
```

If either of the two approximations were calculated alone, the first four lines would still be required although it would be possible to incorporate the calculation of D, which only provides the sign, into the statement in which the approximation is calcualted.

BURR'S APPROXIMATIONS

Burr approached the problem through the use of cumulative moment theory and also derived two approximations.(2) Burr's approximations have the advantage that they are valid for all values of P > 0. They also are conviently expressed in one FORTRAN statement. Burr's less accurate approximation is found in formula 6 and the more accurate in formula 7 and are hereafter referred to as Burr(6) and Burr(7).

Burr(6)
$$z = \frac{\left[(1-P)^{-1/6.158} - 1\right]^{1/4.874} - .644693}{.161984}$$

Burr(7)
$$z = \frac{\left[(1-P)^{-1/6.158} - 1\right]^{1/4.874} - \left[P^{-1/6.158} - 1\right]^{1/4.874}}{.323968}$$

These expressions appear even simpler when written in FORTRAN statments:

```
A = -1./6.158

B = 1./4.874

B6 = (((1.-P)**A-1.)**B - 0.644693)/0.161984

B7 = (((1.-P)**A-1.)**B - (P**A-1.)**B)/0.323968
```

BYARS AND ROSCOE'S APPROXIMATION

It was the objective of these investigators to develop an approximation of the inverse Gaussian which was both more accurate and more efficient than previous transformations. In the early stages of the investigation both polynomial expressions and rational polynomial expressions were considered. The rational expressions showed more promise and in the latter stages of the investigation only such expressions were considered. In the early stages of the investigation a number of variable transformations were considered. It was noted that the transformation $R = P - .5$ resulted in the terms of even degree vanishing from the numerator of the rational expression and in terms of odd degree vanishing from the denominator. This made it possible to obtain an expression containing high powers of R, but with only half the number of terms in numerator and denominator that might be expected.

The coefficients of the rational expression were found by using a least squares approach with successive trials having greater weightings on points at which the approximation was least accurate. The final expression was obtained from the solution of approximately 4300 equations in six unknowns. At that time approximately 3400 of the points considered were between .01 and .03 or between .97 and .99.

The approximation derived is as follows:

for $R = P - .5$ and $0 < P < 1$

$$z = \frac{a_1 R + a_3 R^3 + a_5 R^5}{1 + b_2 R^2 + b_4 R^4 + b_6 R^6}$$

where

$a_1 = 2.505922$	$b_2 = -7.337743$
$a_3 = -15.73223$	$b_4 = 14.97266$
$a_5 = 23.54337$	$b_6 = -6.016088$

This expression is written with the following FORTRAN statements:

```
B = P - 0.5000000

Q = R*R

BR = ((2.505922+(-15.73223+23.54337*Q)*Q)*R)/(1.0+(-7.337743+
&(14.97266-6.016088*Q)*Q)*Q)
```

Note that nested multiplication is used in this expression as well as in the Hastings approximations in order to avoid exponentiation which is an expensive computation.

COMPARISON OF APPROXIMATIONS

The three sets of approximations were compared on two sets of criteria: accuracy and computational efficiency.

Accuracy

The three sets of approximations were used to calculate the values of the z-scores corresponding to each percentile from 1 to 99. These z-scores were then compared to tabled values (3) for mean and maximum deviation on that range.

TABLE 1

MEAN AND MAXIMUM ERROR ON RANGE $0.01 \leq P \leq 0.99$ FOR VARIOUS APPROXIMATIONS

Approximation	Mean Error .01 to .99	Maximum Error .01 to .99
Burr 6	0.00825	0.02206
Burr 7	0.00117	0.00356
Hastings 67	0.00185	0.00279
Hastings 68	0.00028	0.00044
Byars-Roscoe	0.00004	0.00010

The above table shows marked superiority for the Byars-Roscoe approximation on the $0.01 \leq P \leq 0.99$. In the region $0.001 \leq P \leq 0.009$ and $0.991 \leq P \leq 0.999$, the Hastings approximations maintained their approximate mean and maximum error characteristics. The Burr approximations were less accurate in the tails than over the rest of the range, but the Burr 7 approximation was more accurate than either the Burr 6 or the Byars-Roscoe both of which had errors in the tenths place at 0.001 and 0.999.

Computational efficiency

The three sets of approximations were compared for computational efficiency by means of timing a large number of executions of the required FORTRAN statements. The approximations were programmed as shown above. The timing was supplied by an IBM supplied subroutine called INTIME. By calling this subroutine before and after the completion of a DO LOOP which contained the given approximation, it was possible to time to within one one-hundredth of a second the length of time that the DO LOOP executed. Each of the five approximations was placed within a DO LOOP which executed 1000 times and again within a loop which executed 2000 times. For comparison purposes a DO LOOP with no interior statements was also executed 1000 and 2000 times. These loops were executed using both the "fast core" and "slow core" options available on the IBM 360/50 at the Kansas State University Computing Center.

TABLE 2

TIME IN SECONDS FOR EXECUTING VARIOUS APPROXIMATIONS

Approximation	1000 Slow Core	2000 Slow Core	1000 Fast Core	2000 Fast Core
Empty DO LOOP	0.10	0.25	0.03	0.05
Burr 6	5.04	9.91	1.87	3.88
Burr 7	9.49	18.68	3.65	8.03
Hastings 67	2.11	4.78	0.93	1.96
Hastings 68	2.26	4.64	0.97	2.06
Byars-Roscoe	0.47	1.01	0.27	0.53

As can be noted in Table 2, the Byars-Roscoe approximation is faster than the Burr approximations by an order of magnitude and approximately four times as fast as the Hastings approximations. It can further be noted that the Burr 7 approximation takes about twice as long as does the Burr 6 formula. This indicates that the primary computational cost is the additional raising of a real number to a real power. It would further seem that the primary cost in the Hastings formulae is in the preliminary steps since the addition of extra terms increases the cost only slightly.

CONCLUSIONS

On the basis of both accuracy and computational efficiency, the Byars-Roscoe approximation is markedly superior to the approximations provided by Burr and by Hastings on the range from $0.01 \le P \le 0.99$. In all cases in which the scores of interest fall within the given range, that approximation should be used.

Within the range $0.001 \le P \le 0.009$ and $0.991 \le P \le 0.999$, the Hastings 68 formula is superior in accuracy and should be used if a substantial portion of the scores of interest fall within this range and must be accurately transformed. The computational efficiency of the Hastings 68 is sufficiently close to that of the Hastings 67 that the more accurate approximation should always be used.

REFERENCES

(1) Hastings, Cecil, Jr. Approximations for Digital Computers. Princeton University Press, 1955.

(2) Burr, Irving W. "A Useful Approximation to the Normal Distribution Function, with Application to Simulation," Technometrics, 9 No. 4, (Nov. 1967).

(3) Abramowitz, Milton and Stegun, Irene A. A Handbook of Mathematical Functions. Washington, D.C., Superintendent of Documents, U.S. Government Printing Office, 1964.